

# **t**-tests

Week 05

Jennifer Mankin

# Looking Ahead (and Behind)

- So far: Fundamental grammar
  - *Amo, amas, amat...*
  - Next up: Starting to converse!
- This week: *t*-test
- Coming up: Correlation,  $\chi^2$
- Further ahead: The Linear Model

# Take-Away Paper

- You will have some data, which you must use to:
  - Ensure the data contains the necessary measures and values
  - Decide on and create a summary and visualisation
  - Perform an appropriate statistical test
- You should be prepared to **explain** and **justify** your analytic decisions!

## Use Your Resources!

**Everything** we ask you to do in the TAP, you will have practiced multiple times in tutorials, skills labs, and worksheets.

To prep or get help, **come to Skills Labs and practicals!**

# Take-Away Paper

Technicalities:

- Only available for **48 hours** from Monday to Wednesday of Week 7
- You can work on it as long as you want in the time available
- Submit the rendered HTML file on Canvas

Read the TAP Information page on Canvas!

**Important**

THERE IS **NO** LATE SUBMISSION PERIOD!!!

# Other Fun Stuff

## R Meetup

The first R meetup is **this Friday 28th February at 11am** in Pevensey 1 2D10

- Bring snacks/lunch and a computer (if you like), and your enthRsiasm (required)!
- We will give a short, optional intro to website building

## Participants Needed!

- Coding Learning Study (2 SONA credits) One online, one in-person
- Generative AI at University (2 SONA credits) Two online credits
- Lecture Learning Study (NO SONA credits) Enter to win one of ten £10 vouchers

## Awards

### Sussex Awards

Nominate a member of staff who has inspired or made a difference to you

### SavioR Award

Nominate a fellow student on the course who has helped you with R

# Objectives

After this lecture you will understand:

- The concepts behind comparing two means
  - Independent and paired samples  $t$ -tests
- Where the  $t$ -statistic comes from
- How to read histograms and means plots
- How to interpret and report the results of  $t$ -tests

# Comparing Two Means

- Extremely common and fundamental testing paradigm

## The Fundamental Question

Do people in **one group** score, perform, react, behave (etc.) differently than in **another group**?

- Two types (for today!)
  - Independent: different entities/participants in each groups
  - Paired: same entities/participants in both groups
- Very similar logic and interpretation, slightly different maths!

# Taste the Rainbow: Synaesthesia (Redux)

- People with synaesthesia have unusual sensory experiences
  - Experience colours for words, shapes for music, personalities for numbers, etc.

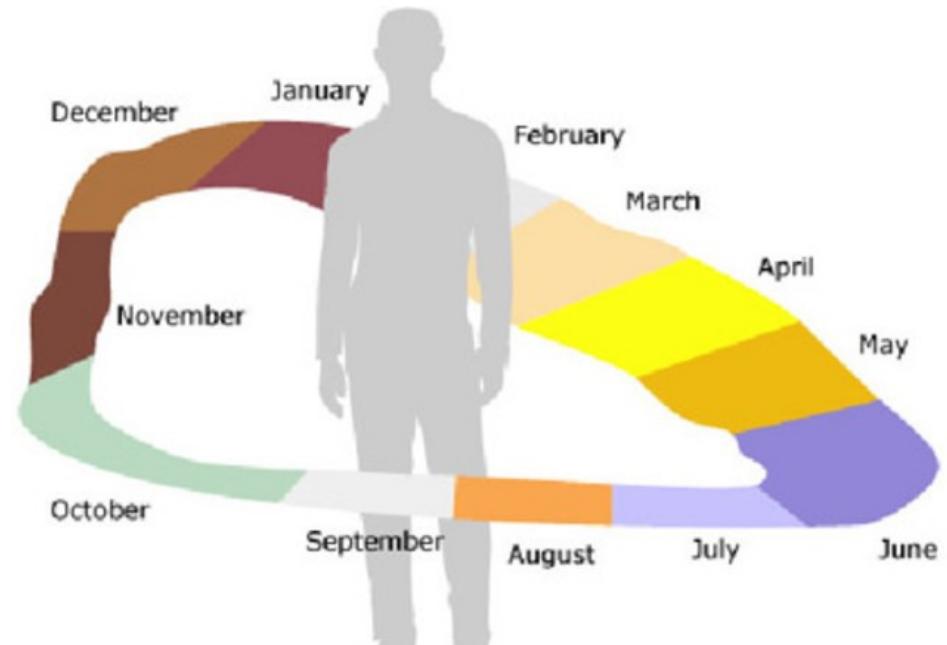
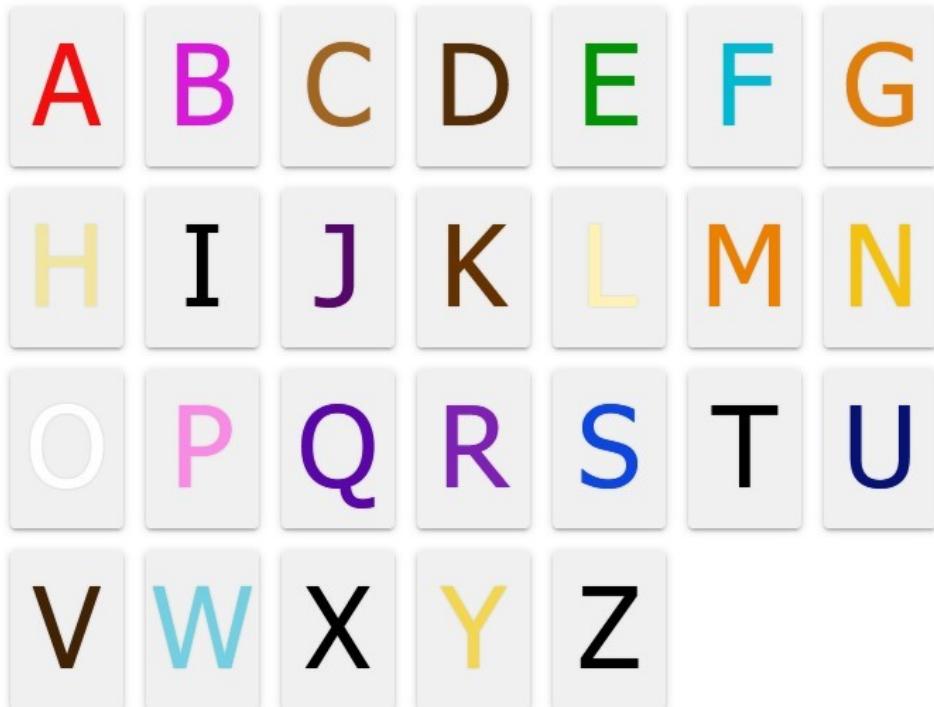


Image Source

# Grapheme-Colour Synaesthesia

- Association between letters/words and particular colours
  - Tends to be consistent throughout life, beginning in childhood
  - So, synaesthetes might tend to notice language/spelling more often

## Research Question

Do synaesthetes have a different cognitive style, compared to non-synaesthetes?

## Conceptual hypothesis

Grapheme-colour synaesthetes have a more language-oriented cognitive style than non-synaesthetes

# Data and Design: SCSQ

- Mealor et al (2016): Sussex Cognitive Styles Questionnaire
  - Includes measures of imagery, **language ability**, and more
  - Example items: “I tend to notice if a word has the same letter repeated in its spelling”; “I enjoy learning new languages”
  - Validated on people with and without synesthesia

## Operational hypothesis

Synaesthetes will, on average, have a **different score** on the Language subscale of the SCSQ than non-synaesthetes

## What is the null hypothesis?

# Summary of Design

## Research Question

Do synaesthetes have a different cognitive style, compared to non-synaesthetes?

## Conceptual hypothesis

Grapheme-colour synaesthetes have a more language-oriented cognitive style than non-synaesthetes

## Operational hypothesis

Grapheme-colour synaesthetes will have, on average, a **different score** on the Language subscale of the SCSQ than non-synaesthetes

## Null hypothesis

Grapheme-colour synaesthetes will have, on average, the **same score** on the Language subscale of the SCSQ as non-synaesthetes

# Having a Look

Let's begin with all scores on the Language subscale together

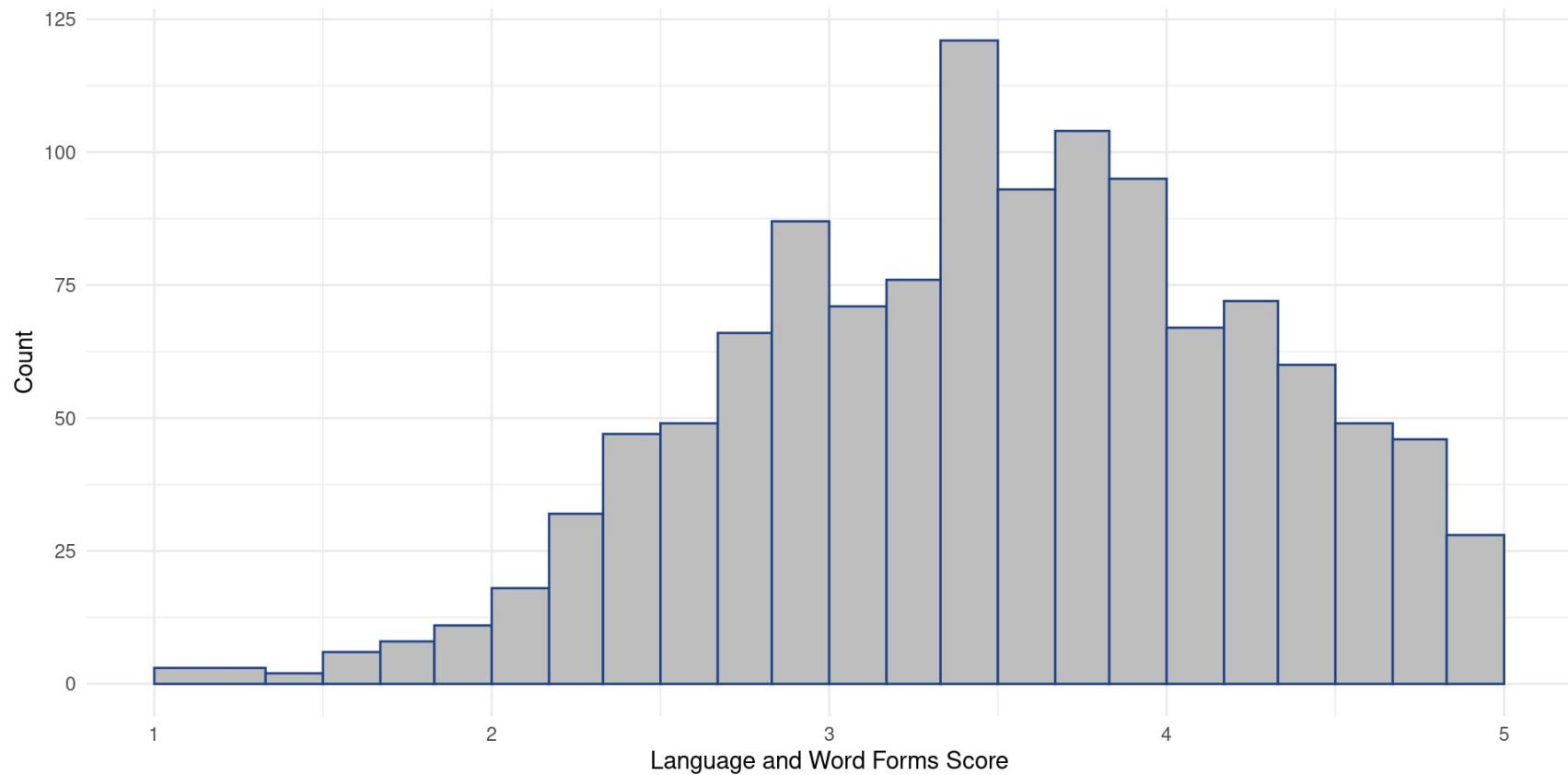


Figure 1: SCSQ Language scores

# Having a Look

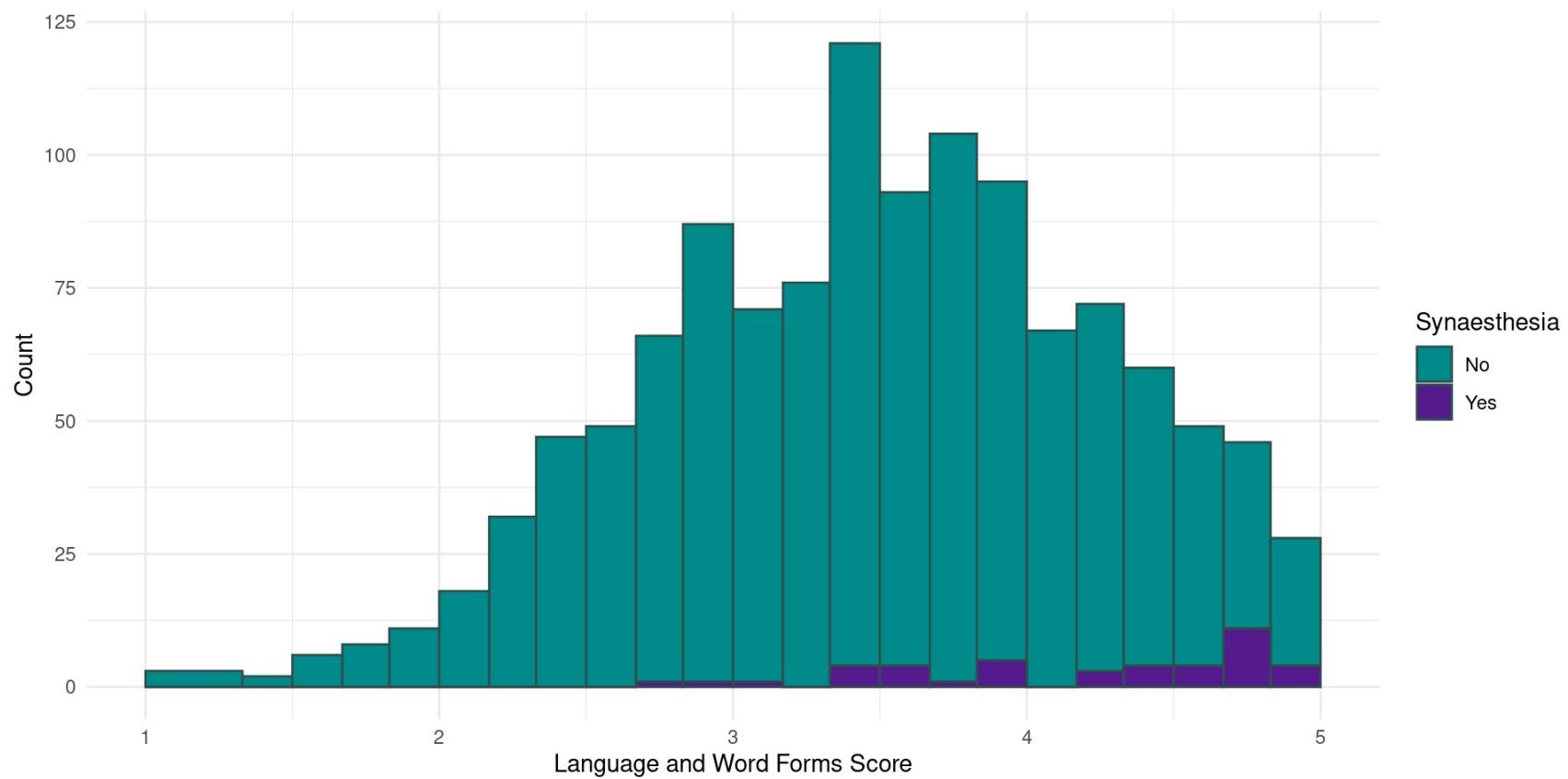


Figure 2: SCSQ Language scores by synaesthesia group

# Having a Look

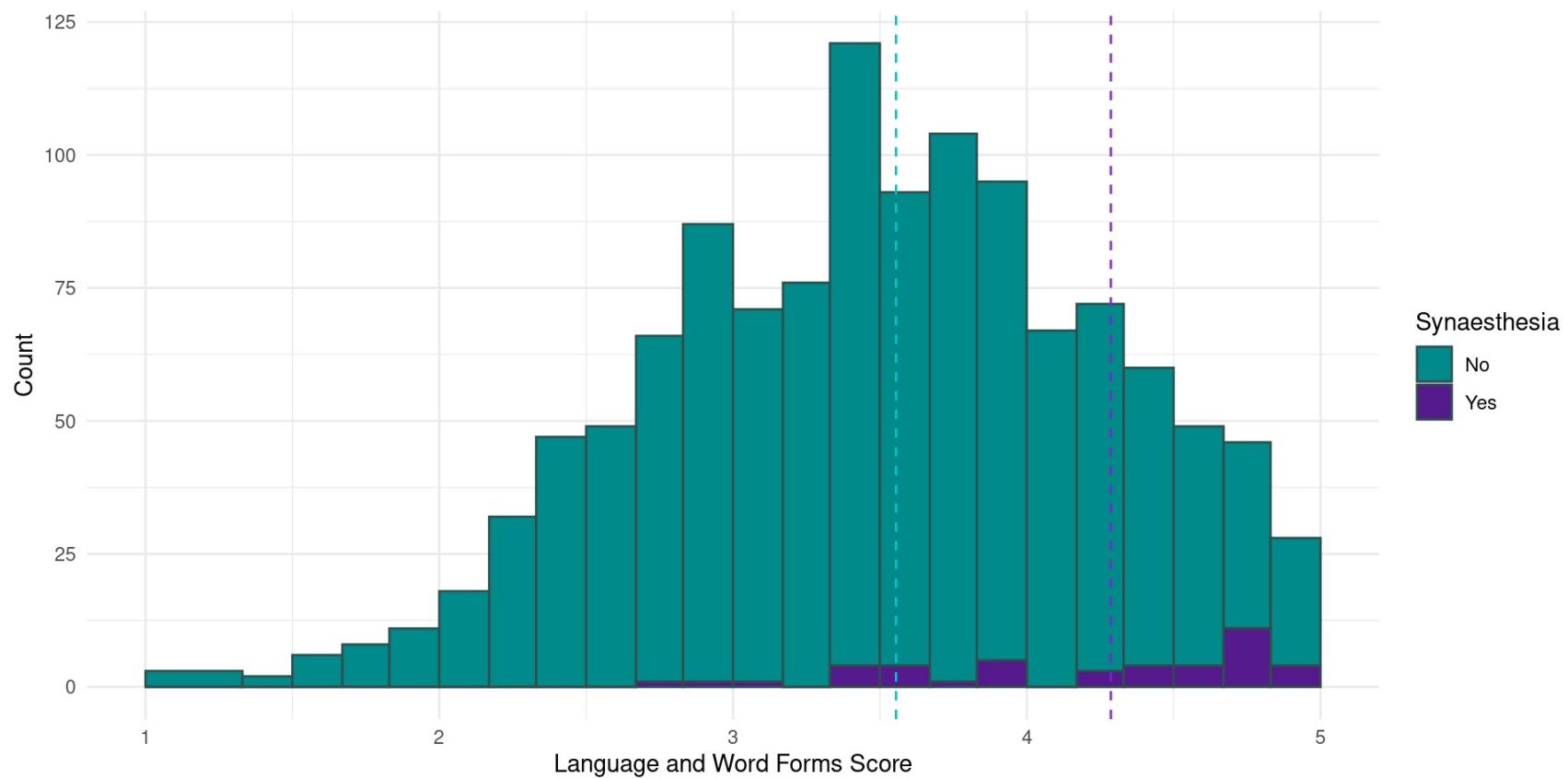


Figure 3: SCSQ Language scores by synaesthesia group, with group means

# Sorted!

- The mean Language score for synaesthetes is higher than for non-synaesthetes
- Are we done?
- Of course not 😊
- How different are these mean scores, *accounting for how much scores vary?*
  - How strong is the *signal* (the difference in means)...
  - Compared to the *noise* (the variation in mean differences)?

# Steps of the Analysis

- Calculate the *test statistic* (signal-to-noise ratio)
- Compare that test statistic to its distribution under the null hypothesis
- Obtain the probability  $p$  of encountering a test statistic of the size we have, or larger, assuming the null hypothesis is true

## Vocabulary: Ratio

A number that captures the relative size of two quantities, expressed as how many times bigger the first quantity is than the second.  
Calculated as the first divided by the second.

e.g., if the seagull-to-human ratio in Brighton is 3, then there are 3 seagulls for every 1 human, or 3 times as many seagulls as humans.



# Calculating the Test Statistic: The Signal

- The “signal” is the relationship of interest
  - The variation in scores explained by group membership
  - Here: the relationship between being a synaesthete (or not) and Language score
- Calculate the mean of each group
- Subtract one mean from the other
- The size of the difference in means is the **signal**

# Having a Closer Look



Figure 4: Mean SCSQ Language scores by synaesthesia group, with 95% CIs

# Calculating the Signal

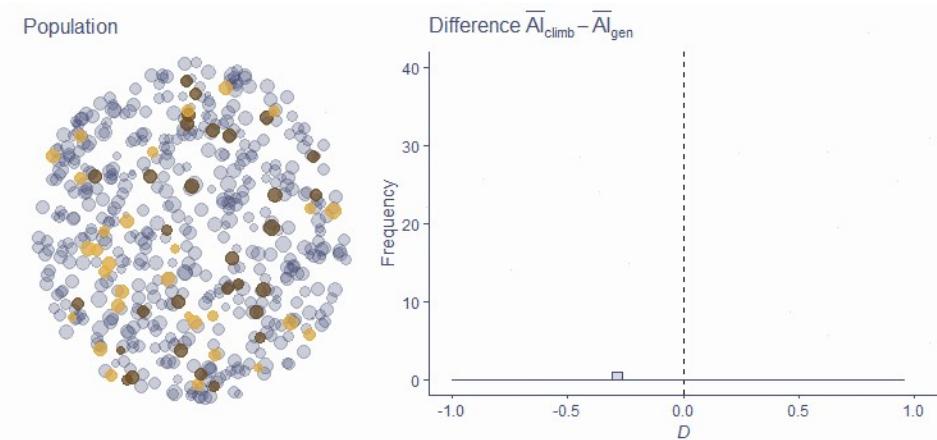
- Let's get some numbers to work with 😊
  - Mean in synaesthete group: 4.29
  - Mean in nonsynaesthete group: 3.55
- Difference in the means (the “signal”) =  $4.29 - 3.55 = 0.74$
- Is this a big difference, compared to how different we might expect any two sample means to be  
**from the same population?**

# Calculating the Test Statistic: The Noise

- The “noise” is the error, the variation NOT explained by group membership
- The differences in means have a sampling distribution!
  - Exactly analogous to the sampling distribution of the mean
- So, the “noise” is the **standard error** of the difference in means between synaesthetes and non-synaesthetes
  - Estimate of how different we expect any two sample means to be **from the same population**

# Side Note: Why “from the same population”?

- Remember: the null hypothesis was that on average, there's no difference between synaesthetes and non-synaesthetes on Language score
  - Basically: group membership doesn't matter
- We're sampling from the same population of scores; synaesthesia is irrelevant
  - Very small differences in means quite likely
  - Very large differences in means quite *unlikely*



# Calculating the Noise

The maths get a bit more complicated here!

- We can derive the standard error of the difference in means from s and N...

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$\hat{SE}_{M_{diff}} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

```
1 syn_lang_syn <- syn_data |>
2 dplyr::filter(syn_graph_col == "Yes"
3 dplyr::pull(scsq_language)
4
5 syn_lang_nonsyn <- syn_data |>
6 dplyr::filter(syn_graph_col == "No")
7 dplyr::pull(scsq_language)
8
9 n_syn <- length(syn_lang_syn)
10 n_nonsyn <- length(syn_lang_nonsyn)
11
12 sd_syn <- sd(syn_lang_syn, na.rm = TRUE)
13 sd_nonsyn <- sd(syn_lang_nonsyn, na.rm = TRUE)
14
15 sd_pooled <- (((n_syn - 1)*(sd_syn^2)) / (n_syn + n_nonsyn - 2))
16
```

[1] 0.1151723

# Steps of the Analysis, Redux

- Calculate the (standardised) *difference* between mean scores
  - Divide the signal (difference in means) = 0.74...
  - By the noise (standard error of the difference in means) = 0.12
- This is our “test statistic”  $t$ , or signal-to-noise ratio

$$t = \frac{\text{signal}}{\text{noise}} = \frac{M_{diff}}{SE_{M_{diff}}} = \frac{0.74}{0.12} = 6.43$$

## Vocabulary: Test Statistic

A number that captures the relationship or comparison of interest. Common examples include  $t, r, \chi^2, F$ . Typically has the general form:

$$\frac{\text{estimate of the relationship or difference}}{\text{estimate of the variation in estimates}} \text{ or } \frac{\text{signal}}{\text{noise}}$$

# We Did It!!

...Or did we?

What does this number mean? How (un)likely is it?

## Steps of the Analysis, Redux

- Compare the *test statistic* to its distribution under the null hypothesis
- Obtain the probability  $p$  of encountering a test statistic of the size we have, or larger, if the null hypothesis is true

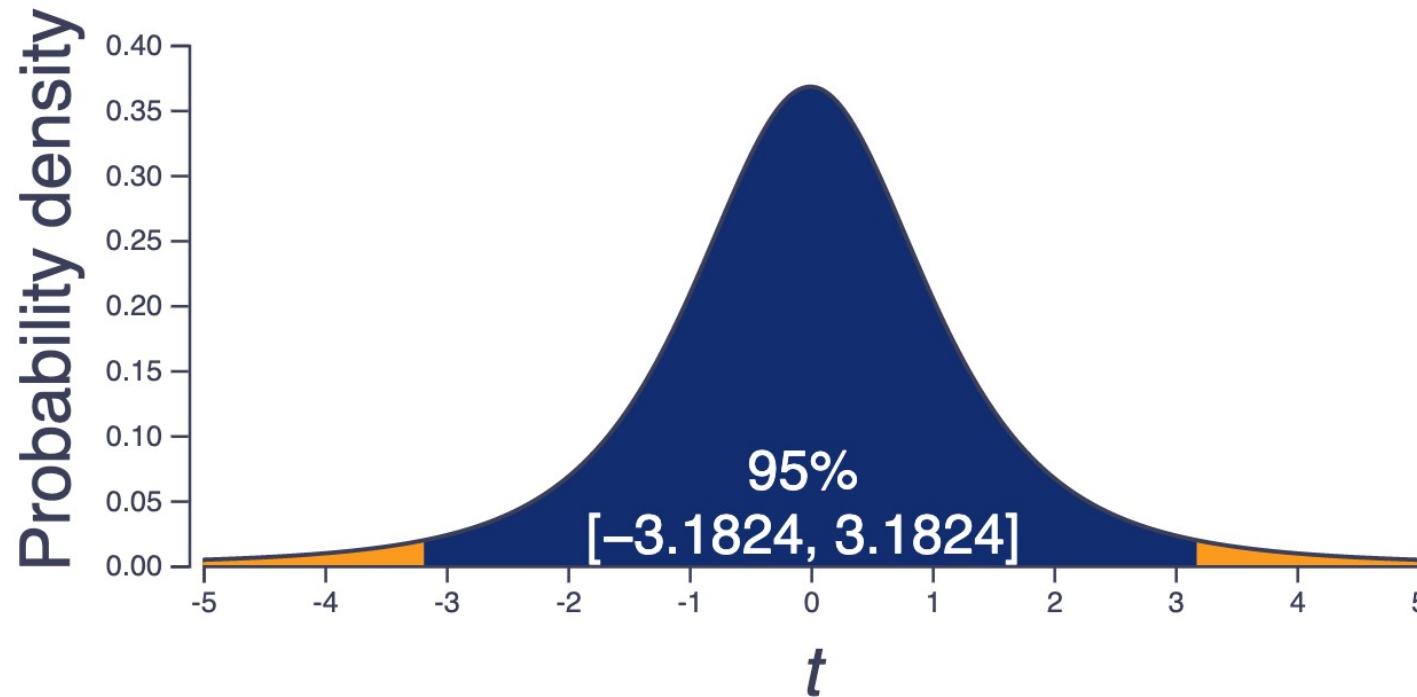
# What's the Point?

- The test statistic  $t$  is the difference in group means divided by their standard error
  - So,  $t$  represents how big the signal is compared to the noise
- Each test statistic has its own **known distribution** under the null hypothesis
  - Known probability of encountering any given test statistic under the null hypothesis
  - Including  $t$ , which is (surprise)  $t$ -distributed

## Important

Larger values of  $t$  are **increasingly unlikely** under the null hypothesis (ie. when the “signal” is, in reality, 0)

# What's the Point?



$df = 3$

# Would You Like Some $t$ ?

- Naturally,  $t$  is  $t$ -distributed (here, with  $N_1 + N_2 - 2$  degrees of freedom)
- What can we conclude, given an  $\alpha$  level of .05?

## If $p > .05$

Our results are *likely* to occur under the null hypothesis

We have **no evidence** that the null hypothesis is *not* true

Conclusion: **RETAIN THE NULL**

## If $p < .05$

Our results are *unlikely* to occur under the null hypothesis

It may in fact be the case that the null hypothesis is *not* true

Conclusion: **REJECT THE NULL**

# Side Note: It's Exactly Backwards

The word “significant” implies

- Big
- Noticeable
- Having a large impact

So we might expect that “significant” means **bigger** than 0.05

It is the *exact opposite of that!*

## Important

Statistical significance expresses surprisingness: the probability of *unlikely* events

So, **small**  $p$ -values are “significant”, not large ones

# That's the *t*

After all of that work...R can do it for us instantaneously!

Two Sample t-test

```
data: scsq_language by syn_graph_col
t = -6.3394, df = 1209, p-value = 3.25e-10
alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
95 percent confidence interval:
-0.9576637 -0.5049972
sample estimates:
mean in group No mean in group Yes
3.554949        4.286279
```

**What do you think?**

What should we conclude?

# Would You Like Some *t*?

## Two Sample *t*-test

```
data: scsq_language by syn_graph_col
t = -6.3394, df = 1209, p-value = 3.25e-10
alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
95 percent confidence interval:
-0.9576637 -0.5049972
sample estimates:
mean in group No mean in group Yes
3.554949      4.286279
```

## Reporting the Results

"On average, grapheme-colour synaesthetes scored higher on the Language subscale of the SCSQ ( $M = 4.29, SD = 0.61$ ) than non-synaesthetes ( $M = 3.55, SD = 0.75$ ). An independent samples *t*-test indicated that this difference was statistically significant ( $t(1209) = -6.34, p < .001, M_{diff} = -0.74, 95\% CI [-0.96, -0.5]$ )."

# Interim Summary

## Independent samples $t$ -tests

- Tests the null hypothesis that two samples come from the same population (i.e.  $M_{\text{diff}} = 0$ )
- Calculate test statistic  $t$ , which expresses signal-to-noise ratio
  - Size of the difference in means divided by the standard error of the difference in means
- Then, evaluate the probability  $p$  of obtaining  $t$  of this size (or larger) under the null hypothesis
- If  $p < \alpha$ , we might conclude that group membership is **associated** with some difference

### No Maths Required

You will not need to memorise, and will not be assessed on, the equations for pooled  $SD$  or  $SE_{M_{\text{diff}}}$ .

# Interlude

## Consider This

Does this constitute evidence that group membership (ie being a synaesthete or not) **CAUSES** a difference in language ability?

- Next up: paired samples  $t$ -test

# Testing Interventions

- Thus far we've seen an **independent samples** test
  - The people in the two groups are **different people**
- Psychology, particularly clinical fields, are often interested in *interventions*
  - That is: does some intervention (therapy, drugs, an activity, etc.) change how people think/feel/ behave etc.?
  - This is a classic **repeated measures** design

# Scenario

You are on a Psychology placement with an NHS service, working with children and their parents. The service provides a “Connections Day” for these families to meet and share their experiences. Before and after the Connections Day, the parents are asked to fill out a questionnaire, which includes a measure of stress.

- Your tasks are:
  - Find out whether the Connections Day event reduces stress in the families who attend
  - Prepare and report the evidence as part of a funding bid to NHS England to continue offering Connection Days

## Note

This scenario is not real, but is inspired by **real NHS placements** and **real tasks** that our current Placement students are doing this year. If this sounds interesting, consider checking out Psychology Placements.

# Summary of Design

## Research Question

Does the Connections Day help parents to reduce stress?

## Conceptual hypothesis

Parents will report reduced levels of stress after the Connections Day, compared to before.

## Operational hypothesis

Parents will have, on average, a **different score** on the stress questionnaire after the Connections Day than before.

## Null hypothesis

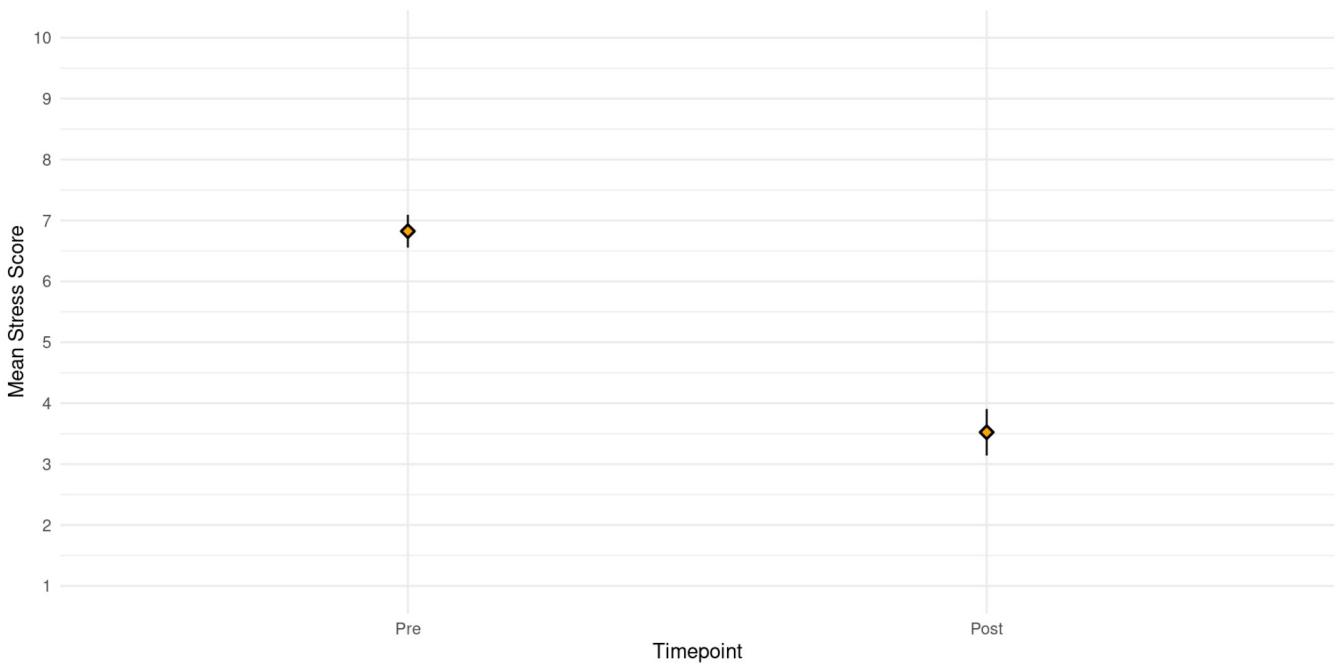
Parents will have, on average, **the same score** on the stress questionnaire after the Connections Day than before.

# Paired (Repeated) Design

- Key difference: the **same people** participate in both conditions
  - The “groups” are no longer groups of people but of *timepoints*: before and after

## Note

This is not real data!



# Example Output

Paired t-test

```
data: rep_data$pre and rep_data$post
t = 21.792, df = 49, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 3.159084 3.800916
sample estimates:
mean difference
      3.48
```

"There was a significant difference in mean reported stress levels from before to after the Connections Day event ( $t(49) = 21.79, p < .001, M_{diff} = 3.48, 95\% CI [3.16, 3.8]$ )."

# Causality

- In our first example, could we conclude that having synaesthesia causes you to pay more attention to language?
- In our second example, could we conclude that attending the Connections Day event *causes* you to have less stress?
- Why is this?

# That's the *t*

- The *t*-test quantifies the size of the difference of two means (signal) compared to the error (noise)
- Independent samples *t*-test
  - Tests means from different entities/participants
  - Independent or “between-subjects” design
- Paired samples *t*-test
  - Tests means from the same entities/participants
  - Repeated or “within-subjects” design
- Establishing causality is a function of **study design** not statistics!

# Have a great day!

## Participants Needed!

- Coding Learning Study (2 SONA credits) One online, one in-person **Finish BEFORE 2pm tomorrow!!**
- Generative AI at University (2 SONA credits) Two online credits
- Lecture Learning Study (NO SONA credits) Enter to win one of ten £10 vouchers

## Awards

### Sussex Awards

Nominate someone for the Sussex Awards here!

### SavioR Award

Nominate someone for the SavioR Award here!

